

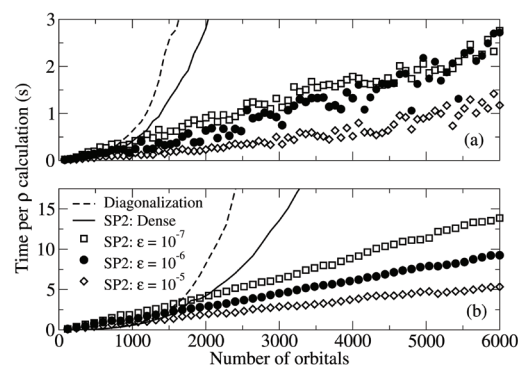
Energy-conserving, Linear-scaling Quantum Molecular Dynamics

Marc J. Cawkwell, Anders M. N. Niklasson, T-1

Molecular dynamics (MD) simulations are used heavily in materials science, chemistry, and biology to study the evolution of structures, defects, and non-equilibrium phenomena at the atomic scale. In an MD simulation atoms move over a number of finite time steps according to the force acting on them. These forces are computed from the interatomic potential that gives the potential energy of the system as a function of the relative positions of all of the atoms. The ability of a simulation to capture the system of interest with high fidelity is determined almost entirely by the physical accuracy of the interatomic potential. It is well established that explicitly quantum mechanical models provide the most accurate descriptions of bonding, but it is not possible to employ these methods in large-scale simulations owing to their prohibitive computational cost.

The scope of quantum-based molecular dynamics (MD) simulations has been limited primarily by the cubic scaling, $O(N^3)$, of the computational cost of solving the quantum eigenvalue problem on the number of atoms, N [1]. Simulations that involve more than a few hundred atoms require a computational cost that scales linearly, $O(N)$, with the number of atoms. Methods have been developed that enable the $O(N)$ scaling of quantum-based potentials that are based on, for example, splitting a large system into a number of non- or weakly interacting smaller ones [1]. The fastest $O(N)$ methods discard long range interactions entirely and enforce “nearsightedness.” While the latter approach is physically intuitive, it leads to errors in the electronic structure and forces that may corrupt an MD trajectory such that the total energy is not conserved. It was not certain that $O(N)$ force computations and energy-conserving MD could ever be reconciled-

Fig. 1. Time per density matrix calculation as a function of the number of orbitals for (a) liquid methane, (b) a polyethylene molecule.



We have developed the capability for performing energy-conserving quantum MD with an $O(N)$ scaling of the computational cost in the LANL-developed code LATTE [2]. LATTE is based on a self-consistent tight-binding model [3], which is the simplest quantum scheme that captures all of the most important interactions in organic molecules—the formation of covalent bonds, charge transfer between species of differing electronegativity, and long range electrostatics. The computation of the energy and forces requires the solution of a Schrödinger-like equation for the density matrix. The computation

of the density matrix is by far the slowest step in the calculation and, furthermore, when traditional approaches are used it as an undesirable $O(N^3)$ scaling. We achieved $O(N)$ scaling for this step by implementing Niklasson's second order spectral projection purification (SP2) algorithm [4] in sparse matrix algebra where only non-zero matrix elements are stored and operated on. By discarding elements whose absolute value is smaller than a threshold, ϵ , at each iteration in the SP2 algorithm, we maintain very high levels of sparsity and enforce nearsightedness. In Fig. 1 we plot the time required to compute the density matrix, ρ , as a function of system size for liquid methane and polyethylene using standard diagonalization techniques, the SP2 algorithm in dense-matrix algebra, and the $O(N)$ implementation of the SP2 algorithm [5]. It is evident that the $O(N)$ scheme becomes rapidly orders of magnitude faster than the $O(N^3)$ methods as the number of atoms increases.

The analysis of the density matrices computed using the $O(N)$ schemes reveals errors that increase as ϵ increases. However, even at the coarsest threshold of $\epsilon = 10^{-5}$ we observe an average error in the interatomic forces of less than 0.02% per atom [5]. Nevertheless, errors of this magnitude are large compared with the numerical precision.

During Born-Oppenheimer MD the atoms following classical trajectories are computed from forces that are evaluated at the self-consistent electronic ground state. Even in the $O(N^3)$ limit these simulations may show a systematic drift in the total energy when the iterations for self-consistency are not converged fully. It was shown recently that this systematic drift can be removed without resorting to full convergence

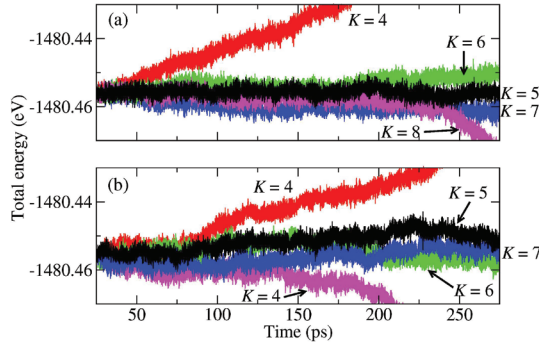


Fig. 2. Total energy versus time for extended Lagrangian Born-Oppenheimer MD simulations with dissipation schemes. (a) Exact forces (b) Approximate forces from $O(N)$ scheme with $\epsilon = 10^{-5}$.

by propagating the electronic degrees of freedom with a geometric integrator. The corresponding extended Lagrangian Born-Oppenheimer MD formalism has been shown to yield stable, energy-conserving trajectories in both first principles and self-consistent tight-binding simulations [6].

The propagation of the electronic degrees of freedom in extended Lagrangian Born-Oppenheimer MD is perfectly lossless, such that errors in the computation of the density matrix accumulate and may eventually corrupt the trajectory. As a result, it is necessary

to introduce dissipation into the propagation to counteract the accumulation of numerical noise [7]. Dissipation in the propagation of the electronic degrees of freedom, $n=\{n_k\}$, which are the starting guesses for the set of partial charges, is given by a Langevin-like dynamics where the modified Verlet integrator at time step i is:

$$n_{i+1} = 2n_i - n_{i-1} + \kappa(q_i - n_i) + \alpha \sum_{k=0}^K c_k n_{i-k} \quad (1)$$

The level of dissipation is controlled by the parameters κ and α , and by the $K+1$ coefficients c_k . In a series of simulations of 50 methane molecules at liquid density and a temperature of 300 K, we investigated the effects of dissipation on the stability of MD trajectories computed with approximate forces obtained from the $O(N)$ scheme with $\epsilon = 10^{-5}$.

These trajectories, which are plotted in Fig. 2, show that the dissipation scheme $K=5$ provides the systematic drift in the total energy of not exceeding $0.05 \mu\text{eV atom}^{-1} \text{ps}^{-1}$ for both the exact and $O(N)$ simulations [5].

The ability of $O(N)$ quantum MD to reproduce the results of otherwise identical simulations performed with exact forces was assessed using a simulation of a $\text{C}_{100}\text{H}_{202}$ polyethylene molecule in the gas phase. The radius of gyration of the molecule computed over 0.55 ns was $16.2 \pm 3.0 \text{ \AA}$ and $16.6 \pm 2.9 \text{ \AA}$ for the exact and $O(N)$ simulations,

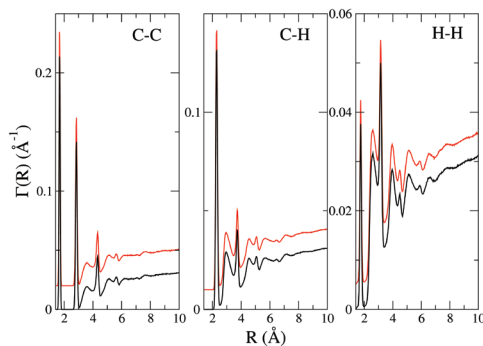
respectively. The average atomic structure of the molecules was assessed using the distribution function:

$$\Gamma_{AB}(R) = \frac{1}{N_A N_B \Delta R} \sum_{i=1}^{N_A} \sum_{j=1}^{N_B} \delta(R_{ij}; R + \Delta R), \quad (2)$$

where A and B label species, N_A and N_B are the numbers of atoms of species A and B, R_{ij} is the distance between atoms i and j , R is the radial distance from atom i , and $\delta(R_{ij}; R + \Delta R) = 1$ if $R \leq R_{ij} < R + \Delta R$ and is zero otherwise [5]. The time-averaged distribution functions for carbon-carbon (C-C), carbon-hydrogen (C-H), and hydrogen-hydrogen (H-H) pairs are plotted in Fig. 3. The $\Gamma(R)$ computed from the exact and $O(N)$ forces are indistinguishable, which strongly suggests that the application of a numerical threshold to enforce nearsightedness in the computation of the density matrix does not affect the ability of the model to capture the salient features of bonding in covalent molecules.

In summary, we have demonstrated for the first time that extended Lagrangian Born-Oppenheimer MD with optimal dissipation enables stable, energy-conserving simulations of extended quantum systems with approximate forces and a low pre-factor $O(N)$ cost. Hence, there is no obstacle to the application of explicitly quantum methods to large-scale simulations in materials science and biology in those situations where empirical potentials are known to be inadequate.

Fig. 3. Time-averaged distribution functions for gas-phase $\text{C}_{100}\text{H}_{202}$. The black and red lines correspond to the simulations computed using exact forces and approximate forces computed using the $O(N)$ scheme with $\epsilon = 10^{-5}$, respectively. The plots have been offset for clarity.



- [1] Goedecker, S., *Rev Mod Phys* **71**, 1085 (1999).
- [2] Sanville, E.J. et al., "LATTE," LANL Technical Report LA-CC 10-004; <http://savannah.nongnu.org/projects/latte> (2010).
- [3] Elstner, M. et al., *Phys Rev B* **58**, 7260 (1998).
- [4] Niklasson, A.M.N., *Phys Rev B* **66**, 155115 (2002).
- [5] Cawkwell, M.J. and A.M.N. Niklasson, *Phys Rev Lett*, to appear (2012).
- [6] Niklasson, A.M.N. et al., *Phys Rev Lett* **97**, 123001 (2006).
- [7] Niklasson, A.M.N. et al., *J Chem Phys* **130**, 214109 (2009).

Funding Acknowledgement

LANL Laboratory Directed Research and Development Program